양준영

포트폴리오

Backend · LLM/RAG · AI 기능 개발 · 실 서비스 운영 경험

부산대학교 정보컴퓨터공학부 학사 재학 / 2021년 3월 ~ 2027년 2월 (예정)
부산대학교 AI융합교육원 학생연구원 / 2026년 3월 ~ 현재

y.jun0.711@gmail.com github.com/Neibce jun0.dev/blog 010-9055-5794

PROFILE

어릴 때부터 일상 속 불편을 줄이는 작은 도구를 만들며 소프트웨어에 관심을 가져왔습니다. 중학교 시절 시스템 자동 종료 유틸리티를 개발해 Naver Software에 등록했고, 이후 학교 급식 알림 챗봇, 학생 편의 앱 등을 개발하며 실제 사용 흐름이 있는 서비스를 만드는 데 관심을 이어 왔습니다.

최근에는 실사용자가 있는 서비스에서 기능 개발과 운영 이슈 분석을 함께 경험했습니다. 부산대학교 AI융합교육원에서는 약 800명이 사용하는 알고리즘 학습 플랫폼 Code Place의 운영 이슈를 분석하고, vLLM 기반 AI 조교 개발에 참여했습니다. DocDoc에서는 RAG, reranker, SSE 기반 의료 논문 탐색 파이프라인을 구현했고, UniSchedule에서는 Redis Pub/Sub과 AWS ECS/Fargate 기반 분산 백엔드 구조를 설계하고 JMeter로 부하를 검증했습니다.

주요 경험

사용자 800명 플랫폼 운영

부산대학교 공식 알고리즘 학습 플랫폼 운영, K3s 운영 이슈 분석/복구, 기능 추가 및 개선, 클라우드 배포, 부하 테스트 경험

LLM/RAG 기능 개발

로컬 LLM 기반 알고리즘 학습 도우미 개발
vLLM 서빙, RAG, Reranker, SSE 구현

AI/SW 대회 9회 수상

외부 6회(대학총장상 2회) / 교내 3회
아주대 딥러닝 챌린지 1위(총장상)
카카오테크캠퍼스 최종 프로젝트 우수상 등

기술 스택

Backend	Java, Spring Boot, Spring Data JPA
LLM / RAG	RAG, LangChain, Pinecone, Reranking, vLLM, SSE
Databases / Caching	MySQL, PostgreSQL, Redis (Pub/Sub)
Infrastructure	AWS (EC2, ECS/Fargate, S3, CloudFront), Docker (Compose), Nginx, Kubernetes (k3s)
Tools / Observability	JMeter, Git/GitHub, Swagger/OpenAPI, GitHub Actions, Grafana, Prometheus, Sentry, k9s
Agentic AI	현재: Codex (Goal, Multi Agent, MCP, Skills), Copilot, Gemini Code Review / 이전: Cursor, Antigravity

주요 활동 및 경력

2026.03 - 현재
부산대학교 AI융합교육원
산학협력단: 학생연구원
코드플레이스 기능 개발, K3s 운영 이슈 분석, vLLM 기반 AI 조교 개발, 서빙

2025.03 - 2026.02
APPTIVE
교내 서비스 개발 동아리: 백엔드 멘토 / 개발
신규 백엔드 부원 멘토링, OnGi 프로젝트 팀장, 최우수상 수상, 기획/디자인/프론트엔드 간 조율

2025.04 - 2025.11
카카오테크캠퍼스 3기
부트캠프: 교육생 / 백엔드 테크 리더
Spring Boot, JPA, 코드 리뷰, 최종 프로젝트 백엔드 테크 리더, 우수상 수상

2024.09 - 2025.08
AID
교내 AI 학술 동아리: CV 스터디 진행
Computer Vision 스터디와 AI 아이디어톤 활동, 아주대 딥러닝 챌린지 개인 1위(총장상) 수상

PORTFOLIO

프로젝트 전체 요약

메인 프로젝트

01 코드플레이스 학생연구원 | 부산대학교 공식 알고리즘 학습 플랫폼

2026년 1분기 - 현재 기능 신규 개발 및 인프라 개선

핵심 기술 k3s · Longhorn · CNPG · vLLM · Grafana · Sentry

핵심 기여 실사용자 약 800명 플랫폼에서 Sentry, Grafana LGTM 기반 관측성 확보, k3s/Longhorn/CNPG 관련 운영 이슈 추적, vLLM 기반 AI 힌트 조교 기능 설계 및 개발

02 DocDoc 팀 | RAG 기반 의료 논문 탐색 보조 서비스

2025년 2분기 RAG 파이프라인 / 백엔드 개발

핵심 기술 LangChain · Pinecone · reranker · SSE · Upstage API

핵심 기여 의료 논문 질의응답을 위해 벡터 검색, reranker 재정렬, SSE 실시간 답변 스트리밍 파이프라인을 구현

03 UniSchedule 팀 | 대학생을 위한 팀 일정 관리 서비스

2025년 3-4분기 백엔드 테크리더

핵심 기술 Spring Boot · WebSocket · Redis Pub/Sub · ECS/Fargate

핵심 기여 Redis Pub/Sub으로 멀티 인스턴스 채팅 메시지 유실 문제 해결, GPT Structured Outputs 기반 시간표 인식, JMeter 부하 검증을 주도, 카카오테크캠퍼스 최종 프로젝트 우수상 수상

04 GRIT 팀 | 실시간 화상 스터디 플랫폼

2025년 4분기 - 현재 LiveKit 구축 및 백엔드 개발

핵심 기술 Spring Boot · LiveKit(SFU) · Redis · OAuth 2.0

핵심 기여 LiveKit SFU 기반 화상 세션 구조와 LiveKit publishData 기반 참여자 간 상호작용 데이터 전달 흐름을 설계/구현, Grafana 대시보드를 통한 LiveKit/백엔드 모니터링 시스템 구축

05 OnGi(온기) 팀 | 가족 건강 관리 및 정서 소통 앱

2025년 2-3분기 팀장 / 백엔드 개발

핵심 기술 Spring Boot · Flutter · FCM/APNs · GCP Cloud Run

핵심 기여 팀장으로 백엔드 API, 가족/건강 데이터 흐름, FCM/APNs 푸시 알림, Cloud Run 배포 자동화를 구현, 기획/디자인/프론트엔드 팀원 간 조율 담당. PNU SW 해커톤 최우수상 수상

추가 프로젝트

01 카카오 선물하기 API 클론 개인

2025년 2-3분기

Spring Boot · JPA · MySQL · OAuth 2.0

상품 조회, 주문, 결제, 인증 흐름을 갖춘 선물하기 백엔드 API입니다. OAuth 인증과 JPA 성능 개선을 중심으로 구현했습니다.

02 BoatTrip 팀 | AI 여행 일정 추천 앱

2025년 2분기 Android / AI 관련 기능 개발

Android Native(Kotlin) · OpenAI API · Web Search · Firebase Firestore · Google Maps API · EfficientNet-Lite

여행지, 일정, 선호 조건을 바탕으로 AI 여행 계획을 추천하는 Android 앱입니다. 지도 연동, 이동 경로 추천, 사진 자동 분류를 구현했습니다.

03 DreamFilm Festival 팀 | 꿈 영화 생성/영화제 개최 서비스

2025년 4분기 백엔드 API / AI 기능 / DB 설계

GPT API · Gemini 3 Pro Image Gen · Database Design

사용자가 적은 꿈을 영화 제목, 포스터, 시나리오로 변환하는 AI 콘텐츠 서비스입니다. 생성 결과 저장과 조회 구조를 설계했습니다.

04 PNU PPlusPlus(피플플) 팀 | 부산대 학생 편의 정보 앱

2024년 3분기 Android 및 CronJob 서버 개발

Android Native(Java) · CronJob · FCM · RSS

부산대 공지, 학사 일정, 식단을 한곳에서 확인하는 Android 앱입니다. 공지 수집, 캐싱, 푸시 알림, Play Store 배포를 담당했습니다.

초기 프로젝트

Nearest-Library-Finder 개인 | 현위치 기반 도서 보유 도서관 조회

2021 HTML · CSS · JavaScript · KolisNet API · Kakao Map API

동천 알리미 개인 | 고등학교 생활 정보 앱

2020 Android (Java) · Node.js (Express) · FCM

PLUTO 팀 | 비주얼 노벨 게임

2018 Python (Ren'Py) · Android (Java) · Google Play Games SDK

해연 급식 알리미 개인 | 카카오톡 급식 조회 챗봇

2017 PHP7 · NEIS · 카카오톡 플러스친구

시스템 종료 예약 개인 | Windows 자동 종료 유틸리티

2015 C# WPF · Visual Studio · ClickOnce · 네이버 Software

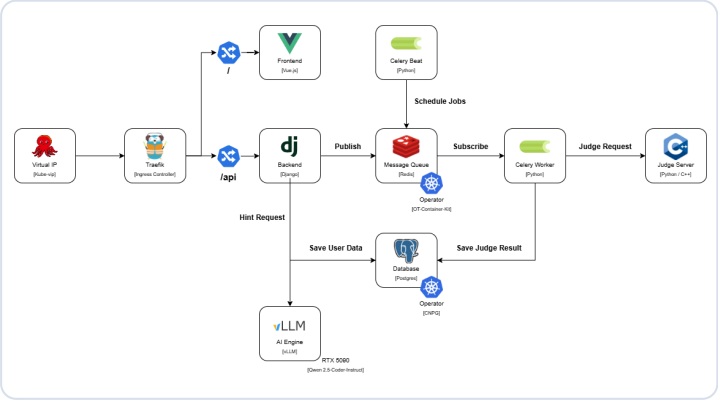
01 코드플레이스

부산대학교 공식 알고리즘 학습 플랫폼

2026.03 - 현재 (개발 진행 중)
운영 서버 - https://code.pusan.ac.kr
https://github.com/pnu-code-place/code-place

개요	코드플레이스는 부산대학교 학생들의 알고리즘 실력 향상을 위해 각종 문제 제공과 채점 서비스를 지원하는 부산대학교 공식 온라인 저지 플랫폼입니다. 약 800명의 사용자가 있으며, 학과 관련 수업의 과제 및 시험, 교내 대회 개최에도 활용되고 있습니다.
역할	부산대학교 AI융합교육원 학생연구원 | 문제 풀이와 학습 경험 개선을 위한 각종 추가 기능 개발, 멀티노드 환경 운영 이슈 분석/해결, 모니터링 환경 개선, vLLM 기반 AI 힌트 조교 개발 및 서빙 환경 구축 수행
기술 스택	vLLM · Qwen · Python/Django · Redis/Celery · Multi-node k3s · Longhorn · CloudNativePG · Sentry · Grafana/Loki/Tempo
주요 기여	• 문제 풀이와 학습 경험 개선을 위한 온라인 저지 플랫폼 추가 기능 개발 • 사용자 코드와 문제 정보를 기반으로 단계별 힌트를 생성하는 AI 힌트 조교 기능 개발 • vLLM 기반 Qwen 모델 서빙 환경 구축 및 AI 힌트 생성 요청 흐름 연동 • Sentry와 Grafana/Loki/Tempo 기반 오류, 로그, 메트릭 확인, Alert 전달 환경 구성 • 운영 가용성 개선 및 Multi-node k3s, Longhorn, CloudNativePG 상태를 같이 확인해 운영 중 이슈 원인 분석 / 복구
핵심 성과	**플랫폼 추가 기능 개발** 문제 풀이, 학습 활동 확인, 관리자 운영 편의 개선 등 온라인 저지 플랫폼의 사용성과 운영 편의를 높이는 추가 기능을 개발했습니다. **AI 힌트 조교 개발/서빙** 사용자 코드와 문제 정보를 기반으로 단계별 힌트를 생성하는 AI 힌트 조교를 개발하고, Qwen 모델을 vLLM으로 서빙하는 환경을 구축했습니다. **운영 이슈 분석 및 복구 대응** 멀티 노드 k3s 환경에서의 CloudNativePG, Longhorn, PVC (Longhorn) 상태와 로그를 함께 확인해 PostgreSQL WAL 급증으로 인한 운영 이슈를 분석하고, 스토리지 확장 및 복구 대응을 수행했습니다.
회고 / 배운 점	코드플레이스를 개발/운영하며 실사용자가 지속적으로 접속하는 서비스에서는 작은 기능 변경이나 짧은 중단도 바로 사용자에게 영향을 줄 수 있다는 점을 체감했습니다. 기능 하나를 배포할 때도 기존 문제 풀이, 채점, 관리자 기능에 영향을 주지 않는지 테스트 서버에서 꼼꼼히 확인해야 했고, 배포 후 오류를 빠르게 감지할 수 있는 구조가 매우 중요했습니다. 또한, 처음으로 온프레미스 멀티노드 환경을 운영하면서 애플리케이션 코드만으로는 서비스 안정성을 보장하기 어렵다는 점을 경험했습니다. 단일 노드나 클라우드와는 다르게 k3s 위에서 Longhorn, CloudNativePG, Redis primary/replica, Celery가 함께 동작하다 보니 장애가 발생하면 원인을 한 계층에서만 찾기 어려웠고, PostgreSQL WAL 누적으로 인한 스토리지 이슈를 분석하며 PVC, Longhorn, CloudNativePG 상태와 로그를 함께 확인해 복구 대응을 수행했습니다. 이 경험을 통해 백엔드 개발자는 기능 구현뿐 아니라 배포 영향 범위, 데이터베이스와 스토리지 상태, 모니터링과 복구 가능성 등 인프라 범위까지 항상 함께 고려해야 한다는 점을 배웠습니다.

서비스 구조도



MAIN PROJECT 01

01 vLLM 기반 AI 힌트 조교 개발

AI 조교 기획 및 Ollama를 vLLM으로 전환해 동시 사용자 50명 처리하기 (블로그)
https://jun0.dev/blog/ollama-to-vllm-serving

1. 문제 상황

코드플레이스에 학생들의 학습에 도움을 주기 위해 AI 조교 기능 도입이 결정되었습니다. 사용자 코드, 문제 설명, 채점 결과를 함께 활용해야 했고, 정답을 제공하지 않고 단계적으로 힌트를 제공하여야 했습니다.
또, 외부 LLM(API)는 비용 문제로 제한되었기에 여러 사용자의 동시 요청을 처리할 수 있는 로컬 모델 서빙 구조가 필요했습니다.

2. Ollama 기반 초기 검증

- Ollama 기반 로컬 LLM 서빙 환경을 구성해 AI 힌트 생성 가능성 검증
- Locust를 활용해 동시 요청 상황에서의 응답 시간과 처리량 테스트
- 테스트 결과, 동시 요청 증가 시 단순 Ollama 로컬 서빙 방식으로는 안정적인 응답 제공이 어렵다고 판단

3. vLLM 전환 및 구현

- KV cache, 요청 배치 처리, GPU 메모리 효율을 고려해 vLLM 기반 서빙 구조로 전환하여 최대 50명이 동시 요청 가능한 환경 구현
- NVIDIA device plugin을 k3s 클러스터에 구성하고, vLLM 워크로드가 GPU 리소스를 할당받아 실행되도록 설정
- Qwen 모델을 vLLM 기반으로 서빙하는 환경 구축
- 사용자 코드, 문제 설명, 채점 결과를 바탕으로 힌트 생성 입력 구성
- 정답을 직접 제공하지 않도록 단계별 힌트 형식과 프롬프트 설계

4. 결과

- 문제 풀이 화면에서 AI 힌트를 요청하고 확인할 수 있도록 기능을 연동
- Grafana 대시보드를 추가해 AI 힌트 요청 상태, 운영 지표 모니터링
- 설문조사 기반 사용자 피드백을 반영해 힌트 방식과 프롬프트를 개선
- 학생들로부터 문제 풀이와 학습에 도움이 되었다는 긍정적인 피드백 확인

02 운영 중 인프라 이슈 분석/해결

WAL 폭주 이후 Longhorn/CNPG 장애 복구기 (블로그)
https://jun0.dev/blog/longhorn-cnpg-troubleshooting

1. 문제 상황

운영 중 DB 스토리지 용량이 고갈되면서 CloudNativePG로 관리되는 PostgreSQL Pod들이 정상적으로 기동하지 못하는 장애가 발생했습니다. DB Pod가 모두 내려가면서 로그인, 문제 풀이, 채점, 관리자 기능 등 모든 기능에 직접적인 영향이 발생하고 있는 상황이었습니다.
코드플레이스는 수업, 과제, 시험, 교내 대회에 활용되는 서비스였기 때문에 빠른 원인 파악과 복구가 필요했습니다.

2. 운영 환경

- 온프레미스 multi-node k3s 환경에서 코드플레이스 운영
- CNPG PostgreSQL 클러스터, Longhorn 기반 분산 스토리지 사용
- Django 웹 서버, Redis primary/replica, Celery 작업 큐, 채점용 Pod가 함께 동작하는 구조
- DB Pod, PVC, Longhorn 볼륨, 각 노드별 디스크 상태가 서로 맞물려 원인 파악이 복잡한 구조

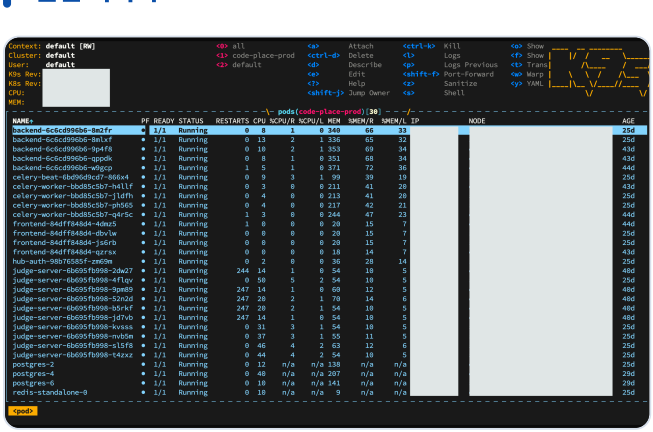
3. 원인 분석

- DB Pod, PVC, Longhorn 볼륨, CloudNativePG 인스턴스 상태를 함께 확인하며 원인 범위 추적
- PostgreSQL WAL 급증으로 DB 스토리지 용량이 고갈되었고, 이로 인해 DB Pod들이 정상 기동하지 못하는 상태임을 확인
- 일부 노드의 Longhorn 여유 공간 부족으로 replica 유지와 볼륨 확장이 함께 제한되는 상황 확인
- CloudNativePG가 일부 PVC를 즉시 복구하지 못하고 Dangling PVC로 처리하는 상태 확인

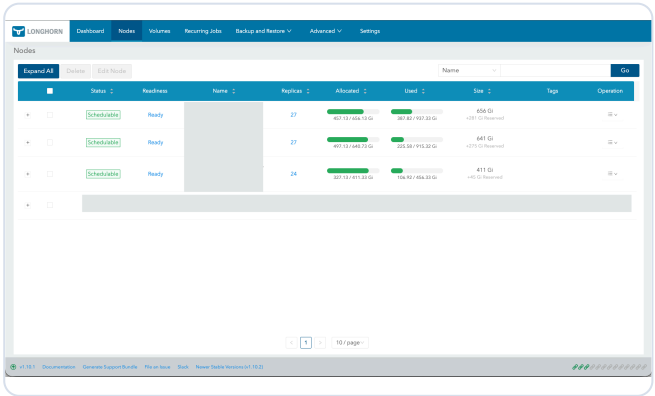
4. 복구 대응

- WAL 누적, PVC 용량, Longhorn replica 배치, CloudNativePG 복구 동작이 함께 얽힌 장애로 판단
- Longhorn에서 용량이 부족한 노드를 일시 배제하고 replica 수를 조정해 최소 복구 가능한 상태 확보
- 스토리지 용량 확장 후 PVC 복구를 진행했으나, CloudNativePG가 일부 DB 인스턴스를 즉시 재구성하지 않는 문제 대응
- DB Pod 재시작과 노드 재부팅을 통해 Mount error 문제 해결
- 추가 디스크를 Longhorn 스토리지 풀에 긴급 편입해 부족한 물리 용량을 확보하고 남은 DB replica 복구

관련 이미지



k9s를 활용한 코드플레이스 운영 리소스 점검



Longhorn 기반 스토리지 상태 점검 및 장애 확인

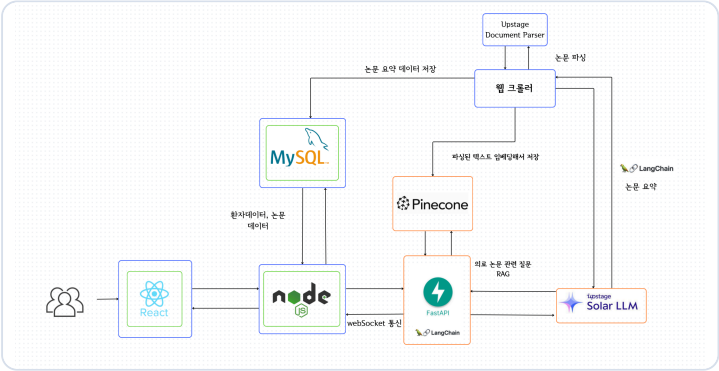
02 DocDoc

의료 취약지 파견 의료진을 위한 RAG 기반 의료 논문 탐색 보조 서비스 (4인 팀)

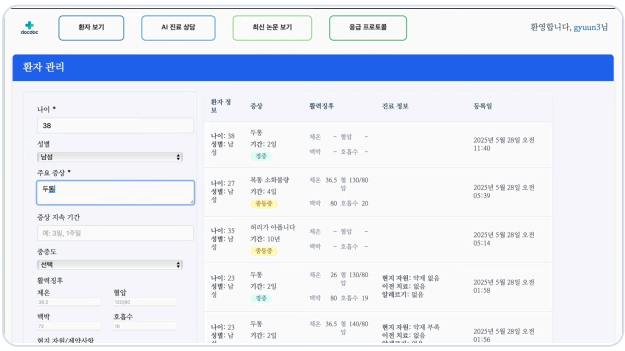
2025.05
PNU x Upstage AI Agent Hackathon 본선 진출
https://github.com/Neibce/DAIC-docdoc

개요	해외 파견이나 의료 접근성이 낮은 환경에서 활동하는 의료진은 제한된 시간과 환경 속에서 최신 의료 논문 기반 근거를 빠르게 찾기 어렵습니다. DocDoc은 수많은 긴 의료 논문들을 검색 가능한 지식으로 변환하고, 사용자의 질문에 맞는 근거 논문을 선별해 답변하는 RAG 기반 의료 논문 탐색 보조 서비스입니다.
역할	RAG 파이프라인 및 백엔드 API 설계/구현 담당
기술 스택	Node.js · LangChain · Pinecone · RAG · Reranking · SSE · Upstage API(Document Parser, LLM) · OpenAI Embedding
주요 구현	• 백엔드 서버와 별도 RAG 서버 간 검색/응답 요청 흐름 구성 • 의료 논문 PDF를 Upstage Parser로 파싱하고, 청킹 및 임베딩(OpenAI API) 후 Pinecone에 저장 • 사용자 질문 기반 벡터 검색 후보를 reranker로 재정렬해 상위 문맥 선별 • SSE로 LLM 생성 과정을 실시간으로 전달해 사용자 답변 대기 부담 완화 • 환자 관리 기능 및 현지 맞춤 의료 가이드 제공
핵심 성과	**의료 논문 데이터 구조화**: 의료 논문 PDF를 파싱, 청킹, 임베딩해 자연어 질문으로 관련 내용을 찾을 수 있도록 했습니다. **검색 후보 재정렬**: 벡터 검색 후보를 reranker로 다시 정렬해 질문 의도와 더 맞는 근거가 앞에 오도록 했습니다. **답변 대기 경험 개선**: LLM 답변이 생성되는 과정을 SSE로 순차 전달해 사용자가 기다리는 부담을 줄였습니다.
회고 / 배운 점	RAG 서비스는 단순 벡터 검색만이 아니라 문서 전처리, 검색 후보 품질, reranker, 응답 전달 방식까지 함께 맞물려야 한다는 점을 배웠습니다. 다만 당시에는 검색 결과와 답변 품질을 주로 직접 확인하며 개선했기 때문에 정량적인 RAG 평가까지 수행하지 못한 점은 아쉬움으로 남았습니다. 이후에는 질문-근거-정답 평가셋을 만들고 검색 정확도와 답변 근거도를 측정해 개선 효과를 정량적으로 검증해보고 싶습니다.

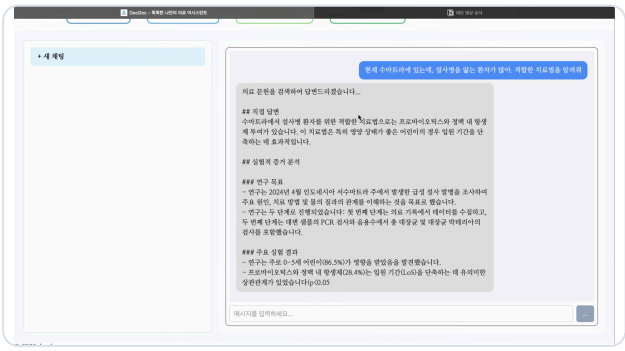
서비스 화면 및 구성도



DocDoc 시스템 구성도



환자 증상, 활력징후, 현지 자원 정보를 관리하는 화면



사용자 질문에 대해 관련 논문 문맥을 검색하고 답변을 생성하는 화면

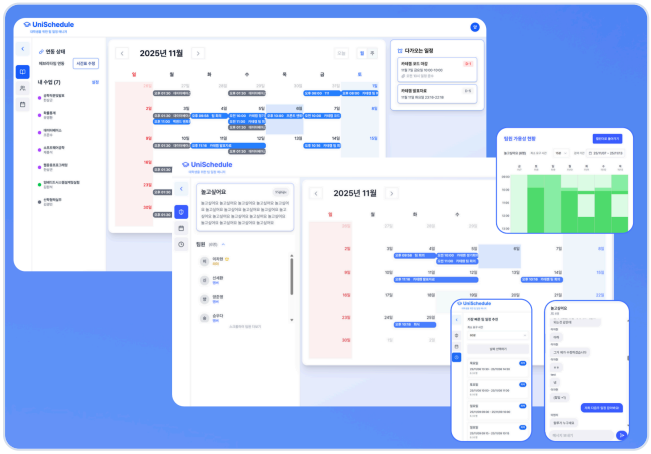
03 UniSchedule

대학생들을 위한 팀 일정 관리 서비스 (6인 팀 - 백엔드 4인, 프론트 2인)

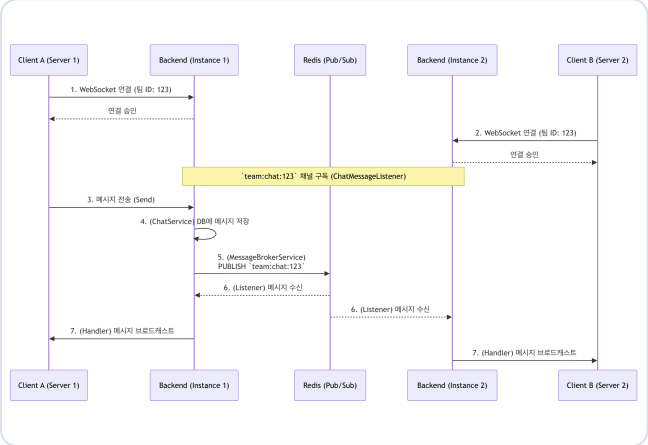
2025.08 - 2025.11
카카오 테크 캠퍼스 3기 최종 프로젝트 우수상 수상
https://github.com/kakao-tech-campus-3rd-step3/Team12_BE
https://github.com/kakao-tech-campus-3rd-step3/Team12_FE

개요	대학생들은 팀 프로젝트나 스터디 일정을 조율할 때 각자의 가능 시간을 단체 채팅방에서 반복해서 확인해야 하는 불편을 겪습니다. UniSchedule은 개인 일정과 팀 일정을 함께 관리하고, 팀원들의 일정 정보를 바탕으로 회의 가능 시간을 빠르게 찾을 수 있도록 돕는 대학생 팀 일정 관리 서비스입니다.
역할	4인 백엔드 팀 테크리더 | 팀원 코드 리뷰, 구현 방향 조율 및 ADR 작성, 프론트 팀과 소통 및 조율, WebSocket 기반 팀 실시간 채팅 구현, JMeter 부하 테스트 주도, GPT 기반 Everytime 시간표 이미지 파싱 기능 구현
기술 스택	Spring Boot · MySQL · Redis Pub/Sub · AWS ECS/Fargate, ALB · JMeter · GPT Structured Outputs
주요 구현	• 4인 백엔드 팀 테크리더로 구현 방향 조율, ADR 작성, 코드 리뷰, API 설계 주도 • AWS ECS/Fargate 배포 환경에서 JMeter 부하 테스트로 오토스케일링 동작 검증 • GPT Structured Outputs 기반 에브리타임(대학생 커뮤니티) 시간표 이미지 파싱 및 변환 기능 구현 • WebSocket, Redis Pub/Sub 기반 실시간 채팅 메시지 전달 구조 적용
핵심 성과	**시간표 등록 부담 감소** GPT Structured Outputs를 활용해 에브리타임 커뮤니티 시간표 이미지를 일정 데이터로 변환했습니다. 사용자가 수업 일정을 직접 입력해야 하는 부담을 줄이고, 초기 일정 등록 흐름을 개선했습니다. **팀 채팅 메시지 전달 구조 개선** 멀티 인스턴스 환경에서 채팅 메시지가 특정 서버에만 머무를 수 있는 문제를 고려해 WebSocket과 Redis Pub/Sub 기반 서버 간 메시지 전달 구조를 적용했습니다. **배포 구조 및 성능 검증** AWS ECS/Fargate 기반 배포 환경에서 JMeter 부하 테스트를 수행해 트래픽 증가 상황에서의 백엔드 동작과 오토스케일링 작동을 검증했습니다.
회고 / 배운 점	UniSchedule에서는 백엔드 테크리더로 참여하며 기능 구현뿐 아니라 팀 개발 흐름을 맞추는 일이 중요하다는 점을 경험했습니다. 여러 기능이 같은 도메인 엔티티를 공유해 PR 간 변경 범위가 겹치는 경우가 있었고, 기능 간 의존성 때문에 작업 순서를 조율해야 하는 상황도 있었습니다. GitHub Flow 기반으로 기능 단위 PR을 나누고, 공통 엔티티 변경은 API 명세와 도메인 책임을 먼저 맞춘 뒤 진행했습니다. 프론트엔드 팀과는 화면 흐름과 API 응답 형태를 조율하며 의견 차이를 정리했고, PR 내 리뷰와 ADR을 통해 구현 방향과 결정 사항을 남겼습니다.

서비스 화면 / 메시지 전달 구조



개인 일정, 팀 일정을 확인하고 팀 회의 가능 시간 조율, 팀 채팅 관련 서비스 화면



멀티 인스턴스 환경에서 WebSocket 메시지를 Redis Pub/Sub으로 전달하는 구조

MAIN PROJECT 03

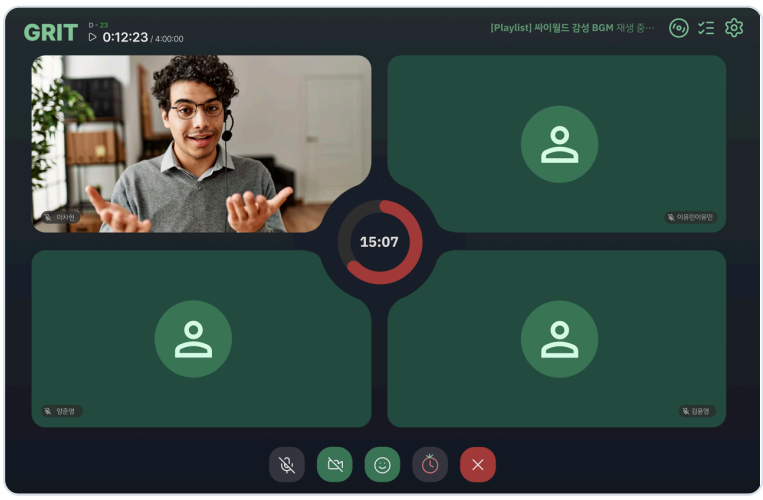
04 GRIT

실시간 화상 스터디 플랫폼 (4인 팀 - 프론트 3인, 백엔드 1인)

2025.12 - 현재 (출시 준비 중)
https://github.com/Don-zo/GRIT

개요	GRIT는 혼자 공부할 때 생기기 쉬운 집중력 저하와 느슨한 학습 흐름을 줄이기 위한 온라인 화상 그룹 스터디 플랫폼입니다. 사용자는 스터디룸에 입장해 친구들과 함께 공부하고, 할 일과 총 학습 시간, 집중 타이머(뽀모도로)를 실시간으로 확인하고 공유할 수 있습니다. 여러 사용자가 함께 참여하는 화상 세션에서 클라이언트 간 직접 연결 부담을 줄이기 위해 LiveKit 기반 SFU 구조를 활용했습니다.
역할	백엔드 개발 담당 | 백엔드 CRUD API, 사용자 가입/인증, 소셜 로그인, LiveKit 서버 관리/연동, 실시간 상호작용 데이터 처리, Redis 기반 스터디룸 상태 캐싱
기술 스택	Spring Boot · Spring Security · OAuth 2.0 · LiveKit · WebRTC · Redis · LiveKit Data packets
주요 구현	• 팀/스터디룸 생성, 프로필 수정 등 핵심 서비스 API 구현 • Spring Security, Google OAuth 2.0 기반 로그인 / 인증 흐름 구현 • LiveKit Access Token 발급, 화상 세션 입장 흐름 구현 • LiveKit Data packets 기반 세션 내 상호작용 데이터 전달 구조 구현 • Redis를 활용해 스터디룸 진행 상태와 방 상태를 캐싱하고 조회 흐름 개선
핵심 성과	**서비스 핵심 API 구현** 팀/스터디룸 생성, 프로필 수정, 사용자 인증 등 서비스 기본 흐름에 필요한 백엔드 API를 구현했습니다. **LiveKit 화상 세션 연동** LiveKit Access Token 발급과 화상 세션 입장 흐름을 구현해 사용자가 스터디룸에서 화상 연결에 안정적으로 참여할 수 있도록 구성했습니다. **스터디룸 상태 및 상호작용 처리** LiveKit API를 매번 호출하지 않도록 Redis를 활용해 스터디룸 상태를 캐싱해, 조회 응답 속도를 개선하고 서버 부하를 줄였습니다. LiveKit Data packets로 세션 내 각종 상호작용 데이터를 전달하는 흐름도 구현했습니다.
회고 / 배운 점	GRIT은 기존 프로젝트들과는 다르게 실시간성이 강한 프로젝트였기에 인증, 팀/스터디룸 API, LiveKit 화상 세션 연동, Redis 기반 캐싱이 하나의 흐름으로 이어지도록 신경써야 했습니다. 프론트엔드 팀원과 화면 동작, 기능 우선순위, API 응답 형태를 자주 조율했고, 그 과정에서 백엔드 구조도 사용자 흐름에 맞춰 여러 번 조정했습니다. 또한, LiveKit API를 자주 호출하면 응답 속도와 서버 부하 측면에서 비효율이 생길 수 있다고 판단해, Redis에 스터디룸 상태를 캐싱했습니다. 이를 통해 실시간 서비스에서는 기존 API 뿐 아니라 자주 조회, 변경되는 내부 상태를 어떻게 관리할지도 함께 신경써야 한다는 점을 경험했습니다.

서비스 화면



스터디룸 서비스 화면: 참여자 카메라, 학습 타이머(뽀모도로), 공부 시간 등을 함께 확인할 수 있습니다.

MAIN PROJECT 04

05 OnGi(온기)

가족 간 건강 관리 및 정서 소통 애플리케이션 (6인 팀 - 기획 1, 디자인 1, 프론트 2, 백엔드 2)

2025.04 - 2025.08
PNU 창의융합 SW 해커톤 최우수상 수상
소개 영상 - https://youtu.be/b4rGJLy9ngk
https://github.com/Neibce/OnGi

개요	OnGi는 부모와 자녀가 떨어져 지내는 상황에서도 건강 상태와 일상을 자연스럽게 공유할 수 있도록 만든 가족 건강, 소통 앱입니다. 부모님은 통증 부위, 복용 약, 운동 시간, 일상 사진 등을 기록하고, 자녀는 이를 확인하며 안부와 건강 상태를 지속적으로 살피고, 같이 일상 사진을 공유할 수도 있습니다. 건강 기록에 그치지 않고 가족 간 정서적 소통까지 이어지는 서비스 흐름을 목표로 개발한 Flutter 기반 크로스 플랫폼 어플리케이션입니다.
역할	팀장 / 백엔드 개발 | 직군 간 요구사항 조율, 구현 범위/우선순위 정리, 가족 건강 기록 등 주요 API, 실시간 푸시 알림 연동, 배포 자동화 중심 구현
기술 스택	Spring Boot · FCM · APNs · GitHub Actions · GCP Cloud Run · Flutter
주요 구현	• 가족 간 초대, 건강 기록(통증 부위, 복용 약, 운동 시간) 작성 / 조회, 일상 사진 등록 등 서비스 주요 API 구현 • 가족별 건강 / 일상 / 점수 및 랭킹 데이터 저장, 접근 권한 설계 • FCM/APNs 기반 푸시 알림 연동으로 앱 사용, 건강 / 일상 기록을 유도할 수 있는 알림 기능 구현 • GitHub Actions와 GCP Cloud Run 기반 CI/CD 파이프라인 구축 및 배포 자동화 • 팀장으로서 개발팀과 기획/디자인 팀 사이의 요구사항을 조율, 소통, 구현 방법/일정 관리 주도
핵심 성과	**가족 건강/일상 기록 흐름 구현** 통증 부위, 복용 약, 운동 시간, 일상 사진 등 가족 건강·일상 데이터를 기록하고 조회하는 API와 접근 권한 흐름을 구현했습니다. **점수·랭킹 기반 참여 유도** 가족별 건강·일상 기록을 점수와 랭킹 데이터로 관리해 사용자가 꾸준히 기록하고 가족 상태를 확인할 수 있는 흐름을 구성했습니다. **푸시 알림 및 배포 자동화** FCM/APNs 기반 푸시 알림으로 기록 참여를 유도하고, GitHub Actions와 GCP Cloud Run 기반 CI/CD 파이프라인을 구축했습니다.
회고 / 배운 점	온기는 다른 프로젝트들과 달리 타 학과 기획/디자인 팀원이 함께한 프로젝트였습니다. 개발 직군끼리 기능을 바로 기획하던 이전 프로젝트들과 달리, "부모님이 부담 없이 기록한다" 같은 사용자 경험 중심의 요구사항을 먼저 이해하고, 이를 API 명세, 데이터 구조, 기술 선택 등 기능적 요구사항으로 구체화하는 과정이 필요했습니다. 또한, 기획/디자인된 내용 중 구현상 제한이 있는 부분에 대해서는 이유를 설명하고, 가능한 대안을 제시하며 조율해야 했습니다. 이 과정에서 개발 용어를 비개발 직군도 이해할 수 있게 풀어 설명하는 능력이 중요했습니다. 팀원이 많아지면서 일정 관리도 더 복잡해졌고, 기능 간 선행 작업과 의존성이 늘어나 팀장으로서 작업 순서와 진행 상황을 더 세심하게 관리해야 했습니다. 타 프로젝트들보다 더 많은 조율과 관리가 필요했지만, 기획/디자인/개발이 각자의 전문성을 살린 덕분에 개발 직군만으로 진행했던 프로젝트보다 훨씬 더 완성도 있는 결과물을 만들 수 있었습니다.

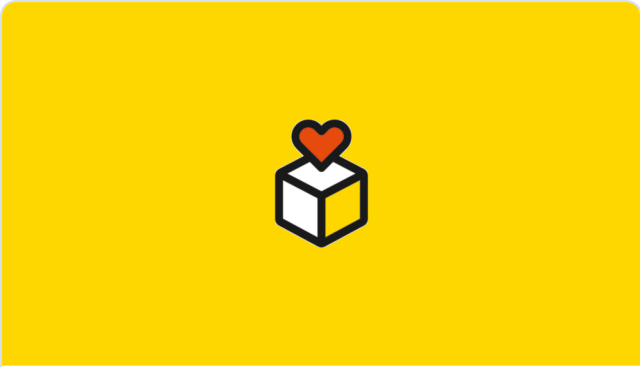
관련 이미지

온기 프로젝트 포스터 | 상세 화면은 소개영상이나 GitHub에서 확인하실 수 있습니다.

MAIN PROJECT 05

추가 프로젝트

메인 프로젝트 외에 최근 진행했던 백엔드, AI 기능, 모바일 서비스 프로젝트들입니다.

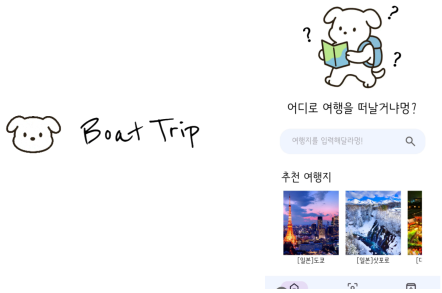


카카오 선물하기 API 클론

2025.06 - 2025.09 / 개인

https://github.com/Neibce/spring-gift-order

- 카카오테크캠퍼스 백엔드 과정에서 상품, 주문, 사용자 인증 흐름을 Spring Boot, JPA, MySQL 기반 REST API로 구현한 선물하기 API 클론 프로젝트입니다.
- Fetch Join과 Batch Size 튜닝으로 JPA N+1 문제를 해결하고 반복 조회 구간의 DB 호출 수를 90% 이상 줄였습니다.
- Kakao OAuth 2.0 로그인 흐름을 구현하고 코드 리뷰를 반영해 서비스 계층과 도메인 책임을 분리하며 SOLID 기반 유지보수성과 확장성을 개선했습니다.

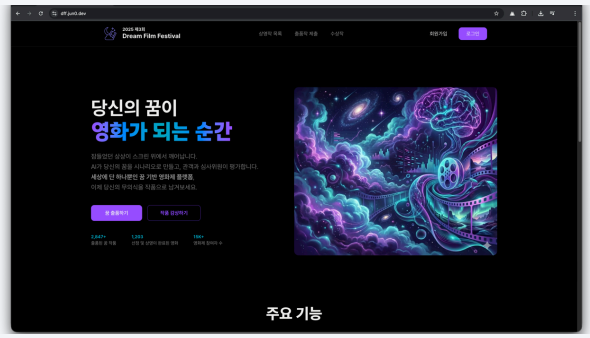


BoatTrip - AI 여행 경로 추천 앱

2025.06 / 팀(2인) · Android 개발, LLM 및 온디바이스 AI 연동

https://github.com/Neibce/BoatTrip

- Android Native(Kotlin), OpenAI API, Web Search, Firebase Firestore, EfficientNet-Lite, Google Maps API 기반 여행 앱입니다.
- 사용자의 여행 조건과 선호도, 검색 기반 여행지 정보를 반영해 LLM 개인화 일정과 이동 경로를 Structured Output으로 생성했습니다.
- 온디바이스 AI 사진 분류, Firestore 데이터 구조, 일정 저장, 캘린더 동기화, 갤러리 연동, Google Maps 경로 표시를 구현했습니다.



DreamFilm Festival - AI 꿈 영화화 서비스

2025.11 / 팀(2인) · 백엔드, AI 플로우, 데이터베이스 설계

https://github.com/Neibce/DreamFilm-Festival

- 사용자가 입력한 꿈을 영화처럼 재구성해 제목, 시나리오, 포스터를 만들고 영화제 형식으로 전시하는 AI 콘텐츠 서비스입니다.
- GPT API로 꿈 서사를 영화 콘텐츠로 확장하고, Gemini 이미지 생성으로 영화 포스터 스타일의 시각적 presentation 경험을 구현했습니다.
- 콘텐츠 저장, 조회, 심사 흐름을 고려해 데이터베이스를 설계하고 인덱스, 뷰, 조인으로 조회 흐름과 관리 편의성을 개선했습니다.

PNU PPlusPlus - 부산대 학생 편의 앱

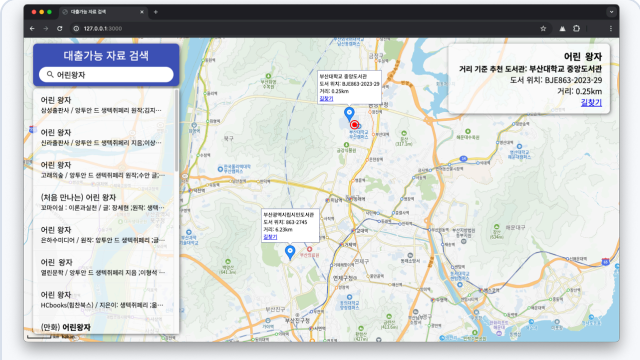
2024년 3분기 / 팀(4인) · 팀장, Android 및 서버 개발

https://github.com/Neibce/PNU-PPlusPlus

- 부산대학교 학생들이 교내 정보를 더 편리하게 확인할 수 있도록 Android 기반 학생 편의 앱을 개발하고 Google Play Store에 배포했습니다.
- CronJob 기반 서버 사이드 크롤링으로 교내 공지사항을 주기적으로 수집하고, 캐싱으로 앱 내 공지 조회 흐름을 개선했습니다.
- FCM 푸시 알림 서버로 주요 공지 알림을 제공하며 모바일 앱, 데이터 수집, 알림을 함께 구현해 사용자 중심 서비스를 개발했습니다.

초기 프로젝트

일상 속 불편을 해결하기 위해 개발했던 초기 개인 프로젝트들입니다.



Nearest-Library-Finder - 근처 도서관 찾기

2021 / 개인

https://github.com/Neibce/Nearest-Library-Finder

- HTML, CSS, JavaScript로 만든 위치 기반 도서 대출 가능 도서관 탐색 서비스입니다.
- 사용자의 현재 위치를 기준으로 선택한 책을 빌릴 수 있는 가까운 도서관을 찾고, KOLIS-NET API와 Kakao Map API로 도서 보유 정보와 위치 정보를 함께 제공했습니다.

Dongcheon Alimi - 고등학교 학생 편의 앱

2020 / 개인

https://github.com/Neibce/DongcheonAlimi

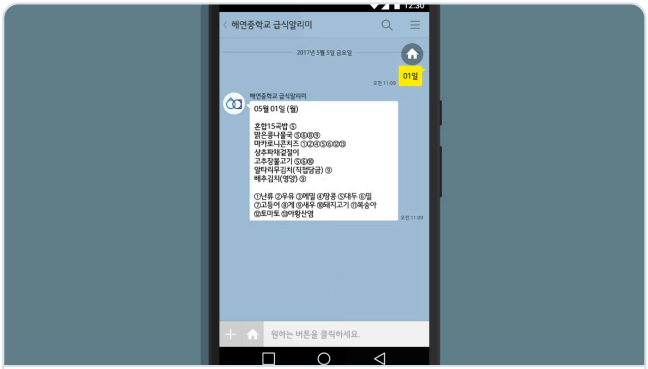
- Android(Java), Node.js(Express), FCM 기반 고등학교 학생 편의 앱입니다. 시간표, 급식, 학사 일정, 공지사항을 한 번에 확인할 수 있도록 만들고 Play Store에 출시했습니다.
- Node.js 서버와 FCM 푸시 알림을 연동해 주요 공지 알림과 건의사항 작성 기능을 구현했으며, 교내 학생 약 50명의 사용 피드백을 반영해 개선했습니다.

PLUTO - 비주얼 노벨 게임

2018 / 팀 · Ren'Py 개발, Android 포팅 담당

https://github.com/Neibce/PLUTO

- Ren'Py 기반 비주얼 노벨 게임 제작에 참여하고 Android 환경에서 실행할 수 있도록 포팅을 담당했습니다.
- Google Play Games SDK를 연동해 모바일 게임 서비스에 필요한 기능을 구현했습니다.



Haeyeon-Lunch-Bot - 급식 알림 챗봇

2017 / 개인

https://github.com/Neibce/Haeyeon-Lunch-Bot

- PHP7과 Simple HTML DOM Parser 기반 NEIS 파싱을 활용한 카카오톡 플러스 친구 급식 알림 챗봇입니다.
- 사용자가 날짜를 선택하면 해당 날짜의 학교 급식 정보를 조회해 반환하도록 구현했습니다.

시스템 종료 예약

2015 / 개인 · C# WPF, Visual Studio, ClickOnce · 네이버 소프트웨어 등록

정해 둔 시간이나 조건에 맞춰 PC를 자동으로 종료하는 Windows용 유틸리티를 개발하고 네이버 Software(현재 서비스 종료)에 배포했습니다.

수상 내역 · 활동 · 자격

수상 내역

2025.11 **카카오테크캠퍼스 최종 프로젝트 우수상 (6인 팀)**
카카오테크캠퍼스 3기
UniSchedule 프로젝트 개발

2025.08 **PNU 창의융합 SW 해커톤 최우수상 (6인 팀)**
부산대학교 정보의생명공학대학장상
OnGi(온기) 프로젝트 개발

2024.11 **PNU CSE 아이디어톤 3등상 (4인 팀)**
부산대학교 정보컴퓨터공학부장상
이미지 인식 기반 페달 오조작 감지 시스템 아이디어 구성

2024.11 **PNU Coding Challenge 장려상 (개인)**
부산대학교 소프트웨어융합교육원장상
부산대학교 알고리즘 대회 개인 참가

2024.08 **제3회 대학 연합 아주 소중한 딥러닝 챌린지 1위 (개인)**
아주대학교 총장상
VLM 기반 zero-shot classification · Public/Private score 1위

2020.09 **2020 부산 코딩경진대회 은상 (개인)**
부산대학교 SW중심대학단장상
부산광역시 주관 알고리즘 대회 개인 참가

2019.10 **부산 SW교육 중등학생 해커톤 최우수상 (4인 팀)**
부산광역시교육연구정보원장상
센서 기반 IoT 절전 전등 개선 버전 개발

2019.07 **POSTECH 무한상상 챌린지 2등상 (4인 팀)**
KT 지속가능경영단장상
의약품 자동 공급 스마트 구급상자 · 임베디드/앱 개발

2019.07 **우송대학교 전국 SW 동아리 경진대회 은상 (4인 팀)**
우송대학교 총장상
센서 기반 IoT 절전 전등 · 임베디드/서버/Android 앱 개발

활동

2026.03 - 현재 **부산대학교 AI융합교육원**
산학협력단: 학생연구원
코드플레이스 기능 개발, K3s 운영 이슈 분석, vLLM 기반 AI 조교 개발, 서빙

2025.03 - 2026.02 **APPTIVE**
교내 서비스 개발 동아리: 백엔드 멘토 / 개발
신규 백엔드 부원 멘토링, OnGi 프로젝트 팀장, 최우수상 수상, 기획/디자인/프론트엔드 간 조율

2024.02 - 2026.01
2021.07 - 2022.04 **IM4U 코딩교육센터**
알고리즘 교육 동아리: 알고리즘 강사
중고등학생 대상 자료구조, 알고리즘, C/C++ 문제 풀이 지도

2025.04 - 2025.11 **카카오테크캠퍼스 3기**
부트캠프: 교육생 / 백엔드 테크 리더
Spring Boot, JPA, 코드 리뷰, 최종 프로젝트 백엔드 테크 리더, 우수상 수상

2024.09 - 2025.08 **AID**
교내 AI 학술 동아리: CV 스터디 진행
Computer Vision 스터디와 AI 아이디어톤 활동, 아주대 딥러닝 챌린지 개인 1위(총장상) 수상

자격 / 어학

SQLD	TOPCIT 610점	TOEIC 755점	네트워크관리사 2급
2025.12	2026.05	2025.03	2021.12

AWARDS · ACTIVITIES · CREDENTIALS